AFRL-HE-WP-TR-2005-0103

AIR FORCE RESEARCH LABORATORY

Developing Concept Learning Capabilities in the COGNET/iGEN Integrative Architecture and Associated Agent-based Modeling and Behavioral Representation (AMBR) Air Traffic Control (ATC) Model

Wayne Zachary
Joan Ryder
James Stokes
Floyd Glenn
Jean-Christophe Le Mentec
Thomas Santarelli

CHI Systems, Inc.
1035 Virginia Drive, Suite 300
Fort Washington, PA 19034

September 2004

Final Report for October 2001 to September 2004

Human Effectiveness Directorate
Warfighter Interface Division
Cognitive Systems Branch
2698 G Street
Wright-Patterson AFB OH 45433-7604

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
September 2004	Final	October 2001 - September 2004

4. TITLE AND SUBTITLE
Developing Concept Learning Capabilities in the COGNET/iGEN Integrative Architecture and Associated Agent-based Modeling and Behavioral Representation (AMBR) Air Traffic Control (ATC) Model

5a. CONTRACT NUMBER
F33615-01-C-6078

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
62202F

6. AUTHOR(S)
Wayne Zachary, Joan Ryder, James Stokes, Floyd Glenn, Jean-Christophe Le Mentec, Thomas Santarelli

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
1710D110

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CHI Systems, Inc.
1035 Virginia Drive, Suite 300
Fort Washington, PA 19034

8. PERFORMING ORGANIZATION REPORT NUMBER
030228.01022

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory, Human Effectiveness Directorate
Warfighter Interface Division
Air Force Materiel Command
Cognitive Systems Branch
Wright-Patterson AFB OH 45433-7604

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-WP-TR-2005-0103

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report describes CHI Systems' efforts during Phase III and IV of the model comparison process of the Agent-based Modeling and Behavioral Representation (AMBR) program conducted by the Air Force Research Laboratory. Phases III and IV focused on modeling and simulating concept learning within the context of a simulated Air Traffic Control (ATC) work environment. CHI Systems extended the COGNET/iGEN framework used in prior AMBR phases, implementing a general capability to learn the conditions under which each of a disjunctive set of goals or actions should be taken. Deeper extensions to the system enabled the representation of memory decay, rehearsal, and proactive interference needed to model human learning performance. The Phase III and IV COGNET/iGEN model execution results were compared with the results of three other models (ACT-R, D-Cog, EPIC-Soar) and results collected from human trials. On most measures (accuracy, response time, workload) iGEN model results were indistinguishable from human performance and, overall, iGEN results provided a better fit to the human data than the other models tested.

15. SUBJECT TERMS Cognitive Architecture, Cognitive Agent, Cognitive Model, Human Behavioral Representation (HBR), Cognitive Modeling, Concept Learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			John L. Camp
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	50	19b. TELEPHONE NUMBER *(include area code)* (937)255-7773

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

THIS PAGE LEFT INTENTIONALLY BLANK

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

This report describes CHI Systems' model development effort as part of Phase III and IV of the model comparison process of the Agent-based Modeling and Behavioral Representation (AMBR) program being conducted by the Air Force Research Laboratory. Phase I dealt with modeling human real-time multi-tasking performance in a simplified ATC task environment built by the fly-off moderator, BBN Technologies (Pew, Tenney, Deutsch, Spector & Benyo, 2000). Within Phase I, the AMBR teams were tasked to create models that could perform the ATC task in a human-like manner, generating performance measures, response times, and post-task subjective workload assessments within ranges documented for human subjects. Phase II dealt with the effects of applying newly adopted distributed simulation standards (IEEE 1516-2000, 1516.1-2000 and 1516.2-2000) to the Phase I models (Tenney & Spector, 2001). Phases III and IV focus on learning, specifically on modeling and simulating concept learning within the context of the existing ATC work environment (see also, Glenn, et al., 2003). Within the phases reported herein, CHI Systems extended the COGNET/iGEN framework used in prior rounds to the problem of concept learning in the AMBR task.

1.1 Concept Learning Problem

There are many different kinds of human learning, ranging from simple classical conditioning of reflexes up through the complex knowledge and skills that take years to develop. For the AMBR program, a problem in concept learning was selected that had previously been extensively studied and that also seems potentially relevant to military applications. This problem was also selected because of the challenge that it poses for models of learning — simple learning models are clearly not capable of predicting some surprising but consistent aspects of human performance in this task. The task, based on an original study by Shepard, Hovland &

Jenkins (1961) and a replication and modeling investigation by Nosofsky et al. (1994), involves asking subjects to learn binary category membership of stimuli that are characterized by binary values on three simultaneous dimensions. In the air traffic control (ATC) context of AMBR, the three dimensions are the size (large or small), and fuel level (high or low) of an aircraft and the atmospheric turbulence around the aircraft. For aircraft characterized by those three dimensional values, the experimental subject must learn which aircraft are to be approved for altitude change requests (the first category) and which are not to be approved for such requests (the second category). With the guarantee of a consistent rule for category membership and only eight possible stimuli to be learned, this might seem to be a very easy learning task. Shepard et al. (1961) noted that the learning task is indeed easy if only one stimulus dimension is relevant (e.g., approve altitude changes for large aircraft only), but that the problem becomes difficult when two or more dimensions are relevant. They noted that the complexity of the logical rules that map between the stimuli and the correct responses could be characterized in terms of just six possible cases. Category 1 (Shepard's Type I), the simplest case, defines the decision rule in terms of just one stimulus dimension, with one value on that dimension indicating the correct response. The learning task for the subject in this case is to determine which dimension is relevant and which value on that dimension signals which response. For the Category 2 (Shepard's Type II) case, the decision rule depends on two dimensions with the third being irrelevant. For Categories 3-6 (Shepard's Types III – VI), the decision rule requires specification of all three dimensions, but with the mapping being somewhat different in each case. The mappings for Categories 3, 4, and 5 can be specified as a Category 1 rule (one dimensional mapping) with two exceptions; the differences between these cases pertain to the relationship between these exceptions and the dimensional structure. The case VI mapping seems to require

the most complex rule specification of all, using all three dimensions or at least 3 exceptions to any simple rule. Shepard et al. (1961) reported experimental data indicating that subjects learned Category 1 rules easily but had extreme difficulty in learning Category 6 rules, with intermediate difficulty for the remaining rules. Nosofsky et al. (1994) replicated the Shepard et al. study and then investigated how several different types of connectionist models could serve to represent the data. They found that a 3-layered, feed-forward neural network model provided the best general fit for this data. AMBR concept learning included stimuli based on Category 1, 3, and 6 rules.

1.2 Overview of Approach and Report

Our basic premise is that concept learning can be characterized as hypothesis testing; that is, an individual posits a classification rule and tests it through experience with cases and feedback, maintaining the hypothesis until there is counterevidence or some other reason to abandon it. At that point the rule is discarded and a new one posited. The fact that the Shepard et al. subjects were able to articulate the rules that they learned in logical form, with varying degrees of efficiency supports such a premise. Following from this assumption, we view concept learning as amenable to a symbolic processing formulation, such as is used within COGNET. The view of concept learning as hypothesis testing also follows from early cognitive psychologists such as Bruner, Goodnow & Austin (1956), and it is consistent with Newell's more recent view (1990) of problem solving as a search process and learning as taking place in service of problem solving. This latter point, is particularly relevant to relating learning to the broader perspective of human performance modeling and simulation and distinguishing it from a purely computational exercise.

Section 2 discusses the modifications to the COGNET/iGEN cognitive architecture to incorporate learning. Sections 3, 4, and 5 describe the model changes and results in Phase III, IV, and IV-resubmit, respectively. Section 6 then provides some conclusions.

2. ADDING LEARNING TO COGNET/IGEN

COGNET is an executable cognitive architecture (see Pew and Mavor, 1998), although unlike most analogous systems it was created specifically for engineering purposes (i.e., as a vehicle for creating practical applications). Originally created as an engine to embed user-models into intelligent interfaces (Zachary, Ryder, Ross & Weiland, 1992), the system has been generalized and extended over time to create a flexible framework for building cognitive agents for use in intelligent training, decision-support, and human performance modeling (Zachary, Ryder, Santarelli & Weiland, 2000). iGEN is an integrated software development environment that supports the authoring, editing, debugging, and integrating of COGNET models (Zachary & Le Mentec, 1999). COGNET/iGEN processing is divided into cognitive, perceptual and action processes, with both perceptual and cognitive processes having access to symbolic memory, which is represented as long-term working memory (LTWM) (Ericcson & Kinsch, 1995) with an adjunct long-term memory. The pre-existing COGNET/iGEN architecture is shown in Figure 1.

Several features of the architecture are particularly relevant to the work reported here:

Emergent attention and multi-tasking. COGNET/iGEN represents attention and multi-tasking via a weak concurrence construct, which allows cognitive attention to focus on only one high-level goal at a time, but to maintain many threads of goal-pursuing activity simultaneously (e.g., as interrupted or pending lines of reasoning). Any high level thread can interrupt any other at any time through a 'pandemonium'-like (Selfridge, 1959) competition process. Competition for the focus of attention is based on context cues,

provided by the changing contents of the LTWM. Thus, attention emerges from the lower level properties of the processing mechanisms, rather than from any explicit attention model, or executive control mechanism.

Metacognition. COGNET/iGEN is unique among cognitive architectures in having *metacognitive* functionality, which provides the system with a symbolic representation of the state of the three primary processing systems via a metacognitive memory. Metacognitive memory includes an internal mechanism called cognitive proprioception. Metacognition also includes representation of pervasive and/or underlying internal states (such as fatigue, or specific beliefs or attitudes) that can affect the way in which goal-oriented processing occurs. The cognitive processor can also execute separate metacognitive processes that use information in metacognitive memory (and other metacognitive knowledge) to affect the execution or application of other procedural or declarative knowledge.

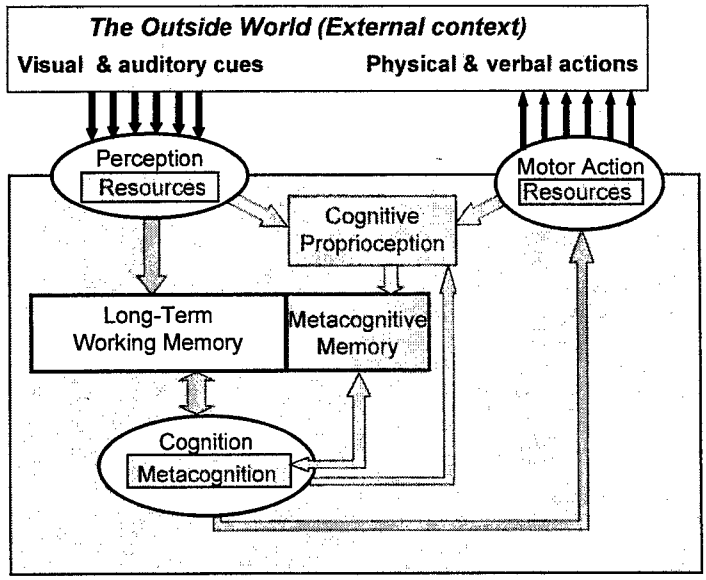

Figure 1. Existing COGNET Architecture

Flexible granularity: The architectural mechanisms in Figure 1 make use of application/domain-specific knowledge and expertise. The representation of expertise is flexible and not tied to any specific level of architectural granularity (e.g., to atomic cognitive processes). COGNET provides different notations to represent declarative knowledge (a blackboard-like representation), procedural knowledge (a GOMS-like conditional goal-subgoal-operator hierarchy notation), and perceptual knowledge (a rule-like notation)

Competence and performance distinction. The core architecture was designed to represent human-like competence -- cognitive abilities unconstrained by internal constraints of time, accuracy, or bandwidth. Human performance is modeled by adding constraining factors of temporality (e.g., processing times), accuracy (e.g., error rates), and/or physical constraints (e.g., vision envelopes, reach limits), through a series of constructs. Limits of sensory or motor mechanisms, for example, are modeled by adding models of specific resources (e.g., eyes, hands, etc.) and their time/accuracy effects.

Creating the capability to learn within the COGENT/iGEN architecture required three types of extensions:

1. Competence level architectural extensions: COGNET/iGEN is a principled system, and the behavior of the component mechanisms and their architectural relationships are governed by explicit principles of operation. The pre-existing system made no allowance for learning of any kind. Thus, to create the competence or underlying ability to learn, a learning mechanism had to be created and integrated into the system. As shown in Figure 2, the mechanism was integrated as a general subcomponent of the cognitive process.

The learning mechanism was created and integrated with the capability to learn the conditions (categories) under which different action options could be taken. Although, in general, learning could apply to any information in memory, the learning mechanism was initially implemented with access to only a subset of long term memory, called the learning memory space, which segmented the previously undifferentiated LTWM into a short-term memory component separate from the remainder of LTWM.

2. Performance level architectural extensions: Additional extensions were required to constrain the behavior of the learning competence to reflect the learning performance of people. In particular, human learning performance is highly constrained by limitations on memory. Thus, it was necessary to add the ability to constrain memory processes by bandwidth and recall limitations which were collectively termed memory moderators.
3. Knowledge representation extensions: The pre-existing representational formalisms did not allow for knowledge components (specifically conditional expressions) which could be learned, nor for goal structures with open (i.e., null) action conditions. By generalizing the construct for pursuing a disjunctive goal set (called 'Decide-Goal'), it was possible to allow the condition sets for each goal in the disjunction to be learned (rather than simply remembered as prior knowledge). The knowledge representation also did not support description of a learning process (e.g., where learning could occur), or metacognitive processes that might control learning. All of these extensions had to be added to COGNET/iGEN as well.

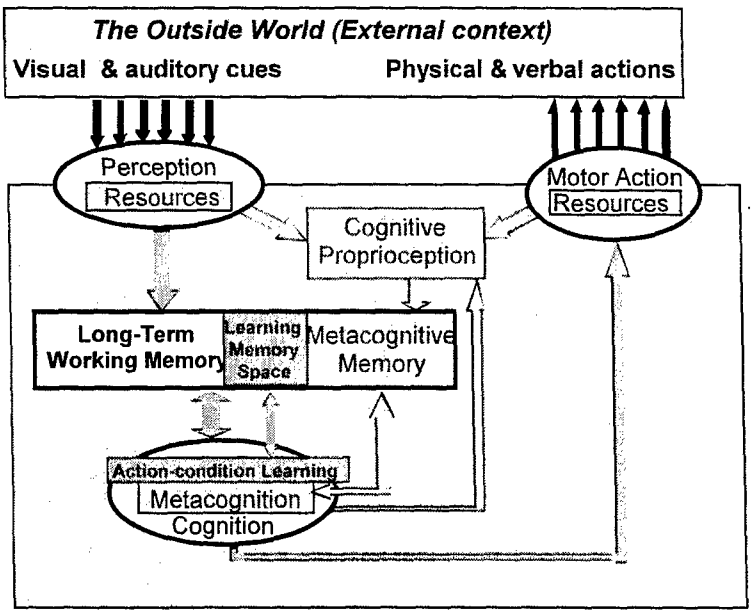

Figure 2. Learning-extended COGNET Architecture

3. PHASE III

The Phase III experimental design was based on manipulation of the following independent variables:

Category (1, 3, 6) – complexity of the decision rules in the learning task

Workload (easy, med, hard) – number of aircraft (0, 12, 16) in the secondary task

These variables result in nine conditions, run as a between-subjects design. Ten subjects were run in each of the nine conditions, for a total of 90 subjects. The model also ran 90 times to provide equivalent performance data. Each run consisted of eight trials (each trial included 16 altitude requests and the relevant number of handoff aircraft for the workload condition). Subjective workload ratings were collected after trials 1, 4, and 8. After the eight trials, the subjects and the model-simulated subjects performed a transfer task.

3.1 Model Development

3.1.1 Approach to Phase III

The Phase III approach built on the existing Phase I/II COGNET/iGEN model. First, we made changes to the Phase I/II model to support the learning task. In parallel, we evaluated learning data and developed a learning strategy for incorporation into the model. We also developed an approach to memory modeling to be incorporated into the learning mechanism. Based on these reviews and analyses, we developed a learning mechanism by extending COGNET/iGEN, and integrated the learning mechanism into the ATC model, retuning and adjusting as needed. Finally, we modified the model to handle the transfer task. Each of these development processes is discussed in a subsequent subsection.

3.1.2 Modifications to ATC task model to support learning task

Changes were made to the Phase I/II model in two areas. First, the model shell was modified to provide communication with the AMBR testbed to support the new Phase III functionality.

Second, the ATC task model incorporated procedures to deal with a new task (the altitude change task). This replaced the speed change task in the Phase I/II model. As shown in Figure 3, the new task includes a goal 'decide_on_response' which invokes the learning component and another goal "update_learning_from_feedback' that processes the feedback from the simulation and provides it to the learning component. In addition, as Figure 4 shows, levels were added to the blackboard panels for keeping track of the altitude-related stimulus attributes and feedback.

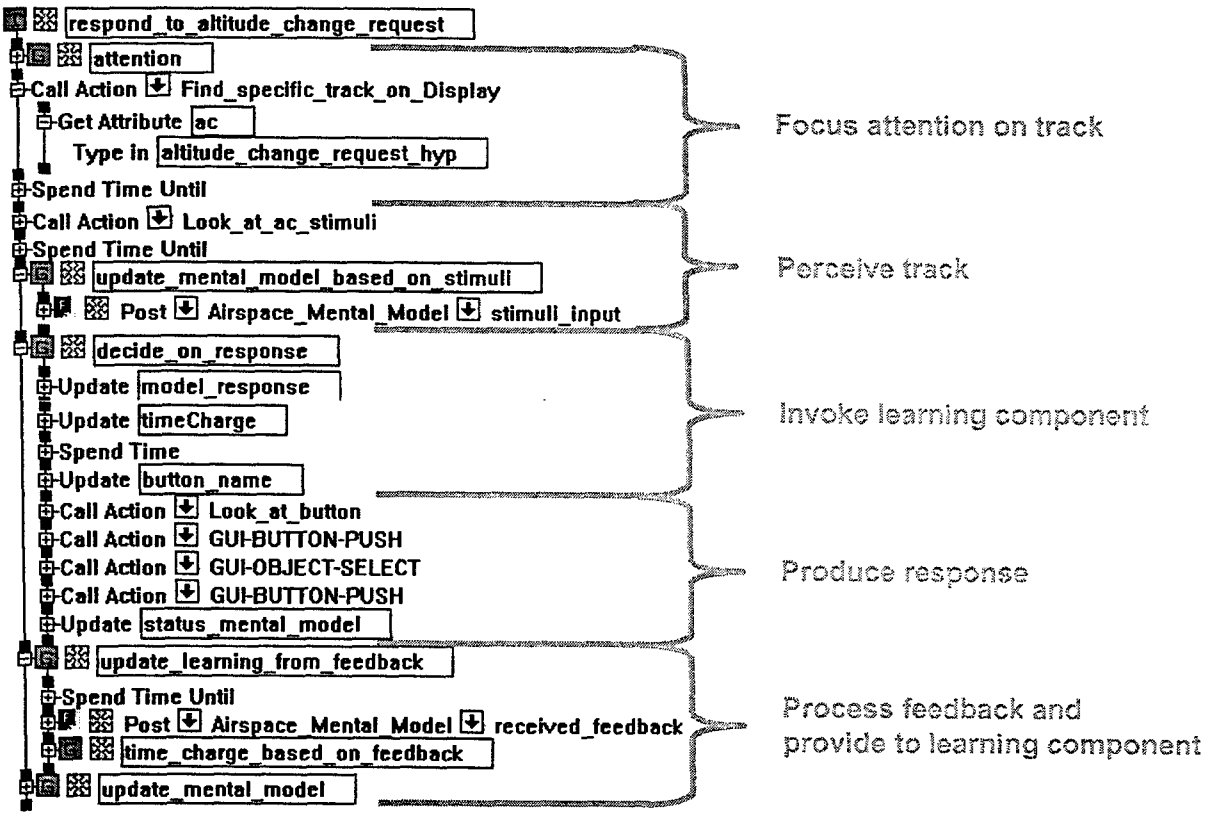

Figure 3. Altitude Request Task with Learning Component Added

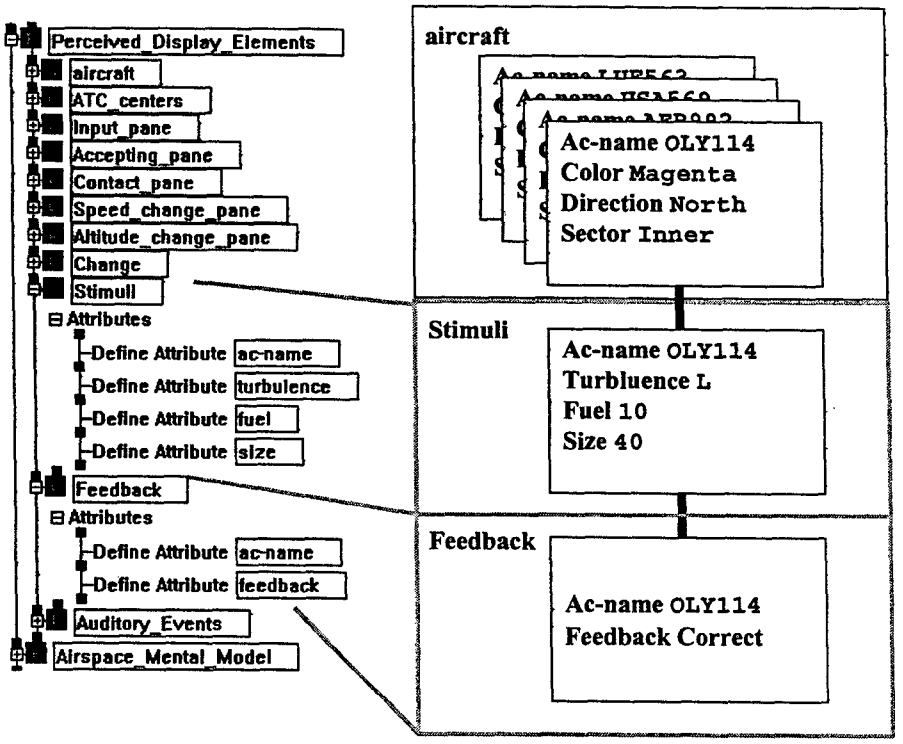

Figure 4. Example Blackboard Additions for Altitude-related Stimulus Attributes and Feedback

3.1.4 Learning Task Strategy

In order to study learning strategies, we developed a simple system for presenting the learning task independently from the ATC task and testbed. Using this learning task presentation software, we conducted a limited cognitive task analysis (CTA), collecting performance data with concurrent verbal strategy self-reporting. Based on this limited CTA in combination with information gathered during human subject experiments as reported by BBN and a review of the category learning literature, we developed a learning task strategy.

Human task performance in category learning is complex, encompassing mixed strategies and encodings. An initial conclusion of our analysis was that the process had to be simplified considerably to model it in a reasonable and parsimonious manner. Some effects that were clear included the fact that learning mechanisms were differentially based on rule complexity, with single-dimension rules being tried first, followed by two-dimension rules, then three dimension rules (three-dimension rules specify unique individual stimulus instances). Subjects' introspection indicates reasoning using partial rules (rules knowing there are exceptions) prior to learning exceptions.

In order to think about and represent category rules, we used a notation in which category rules are represented as a parameter triplet, written in parentheses, specifying a value for each dimension as shown here:

(Size, Turbulence, Fuel Remaining)
(Small, Low Turbulence, 40% Fuel Remaining)
(S,1,40)

Using "x" as a wild card:

(S,x,x) represents a single dimension rule
(S,1,x) represents a two dimension rule
(S,1,40) represents a three dimension rule

A categorization hypothesis consists of one or more Accept-Rules:

(L,x,x)	single one-dimension rule
((L,1,x) (S,3,x))	two partial two-dimension rules

All rules specify request acceptance (at any point in time the current set of rules may contain rules which overlap). Requests are rejected if no rule specifies that they be accepted.

With this scheme, the model would begin with a single one-dimensional guess. If we assume that size is a good first guess (as a primary dimension of interest) the model would begin with (L,x,x) or (S,x,x). Trial results either:

- confirm the rules requiring no change (rules gave correct answer),
- disconfirm one or more rules and require that they be narrowed or removed (rules produced an incorrect "ACCEPT" answer) , or
- are unaccounted for in the current set of rules (rules produced an incorrect "REJECT" answer) and require the addition of a new rule.

In this process, rules are specialized after "ACCEPT" errors, because the rule set is too broad. For example, if a rule (L,x,x) incorrectly accepts the stimulus (L,3,40), the rule would be specialized to one of the form (L,1,x) and ultimately, if appropriate, to the form (L,1,20). In general this process of rule specialization seems to be consistent with observed attempts to find the most powerful rule possible for a given set of stimuli as well as with the need to remember generalities when detailed memory is of necessity incomplete. New rules are generated based on "REJECT" errors, because the rule set is too narrow. For example, after a rule (S,1,x) incorrectly rejects the stimulus (L,3,40), a new rule is generated (L,x,x) yielding two partial rules in the rule set.

From the human perspective, some dimensions are more "important" than others, with reference to the altitude change decision. For example, aircraft size may be the primary consideration for a decision, and turbulence may be a secondary consideration. We refer to this as dimension bias, or a priority among dimensions. In general terms, the initial bias selection would be based on a fixed or random guess, while later bias selection would be based on at least the current stimulus.

The approach described thus far accounts for learning competence. However, learning performance is constrained by memory and other performance constraints. Thus, in addition to the learning *competence*, additional architectural constructs had to be implemented to represent learning *performance*, the most important of which was memory.

3.1.5 Approach to Memory Moderation

A memory moderation model was adapted from the Human Operator Simulator (HOS) work (Glenn, Schwartz & Ross, 1992) and (Lane, Strieb, Glenn & Wherry, 1981). The term memory moderation was used to include various memory-related factors, such as memory load, decay, opportunities for rehearsal, and interference. The model was a two-stage memory model incorporating a short-term memory (STM) with exponential decay and a long-term memory (LTM) with no decay, adapted from the model of Waugh & Norman (1965). Hypotheses are maintained in STM by rehearsal, and converted to LTM with a probability based on memory load and integration decay.

On each trial, the subject attempts to retain some number (1 to 4) of hypotheses (partial rules) in memory. For simplicity, we assume that each stimulus (and feedback) occurs at a regular time interval T. During that time, the subject rehearses the rules to maintain them in the STM and each rehearsal provides an opportunity for conversion to LTM. We designate the

number of slots (non wild-card values) for a particular partial rule as n (n varies from 1 to 3). The memory load is based on the total number of dimension values specified in all rules being remembered (varies from 1-12, since up to 4 rules can be retained and each can have up to 3 values). There is no decay from LTM, but the hypothesis can be fully purged from both LTM and STM by volitional deletion upon determination that the hypothesis is inconsistent with the outcome of a trial. The probability of recalling a hypothesis from STM from one trial to the next is determined by the number of 'slots' that constitute the hypothesis and the amount of rehearsal that is permitted by the memory load of the total stack of hypotheses being maintained. The probability of recall for a rule with n slots is P(short-term) = p^n where p is a learning parameter. N is the total number of slots for all the partial rules. M is the total number of slots that can be rehearsed in the time interval between two stimulus presentations. Note if t is the time to rehearse an individual slot, then M = T/t. Each rule can be rehearsed on average once every N slots, or every Nt seconds. The probability that a rule can be rehearsed after each such interval is:

$$PR(N \text{ slot interval}) = p^n e^{-aN}$$

where a is a decay rate constant.

On average, there will be M/N such rehearsals and recall attempts for each rule between each stimulus, so the probability of recall across the interval between 2 stimuli is:

$$Pr(\text{inter-stimuli}) = (p^n e^{-aNt})^{M/N}$$

and the probability of forgetting is

$$Pf(\text{inter-stimuli}) = 1-(p^n e^{-aNt})^{M/N}$$

We define q as the probability of a rule transferring to long-term memory on each rehearsal (and also upon original formulation). The probability of conversion to long-term between two feedback events is then

$$Pc(\text{long-term}) = 1 - (1 - q)^{M/N} \quad \text{or} \quad Pc(\text{long-term}) = 1 - r^{1/N} \text{ where } r = (1 - q)^M$$

This approach to memory moderation allows problem types to be distinguished by number of opportunities for easily-learned solutions - one dimension rules, or two dimension rules in combination, thus providing performance as expected from humans.

3.1.6 Learning Mechanism

The learning task strategy and memory model described above were integrated to produce a computational learning mechanism within the COGNET/iGEN cognitive architecture. The learning mechanism was developed as a subcomponent of the cognitive system. It has a simple interface to the cognitive system: it receives stimuli, provides a classification in return, and then, receives a positive or negative feedback on its classification. The learning mechanism contains several components (see Figure 5):

- learning memory, divided into short term and long term that contains partial classification rules;
- memory moderator that memorizes or forgets partial rules;
- learning engine that uses the current partial rules to generate classifications and new rules, or to modify existing ones after receiving a feedback;
- metacognitive strategy that influences the learning engine.

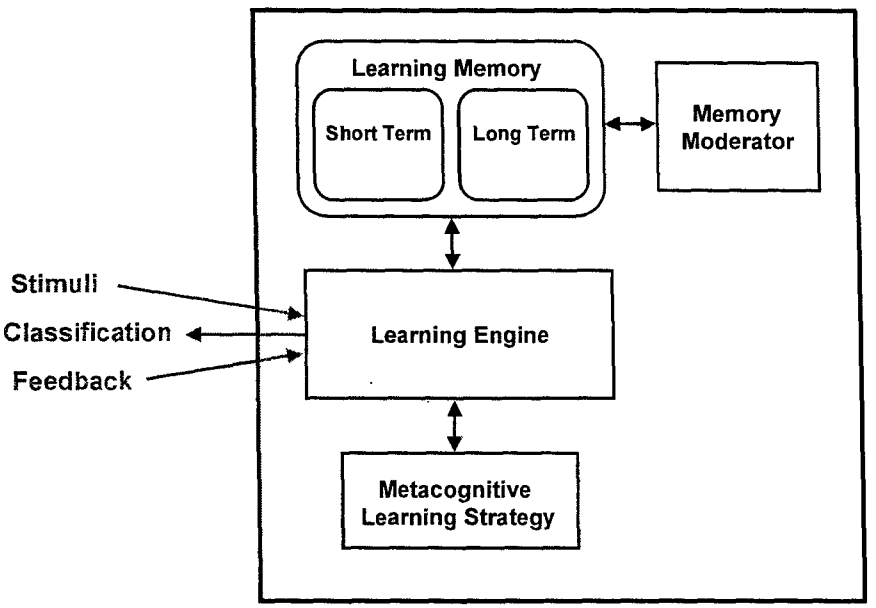

Figure 5. The COGNET/iGEN Learning Mechanism

3.1.6.1 Partial Classification Rules

The knowledge used for classification is represented by a set of partial rules. A rule represents a triplet of stimulus values for each of the three dimensions that correspond to an acceptance case. For example the rule (S,1,40) indicates that the result of the classification must be accepted if size = S (as opposed to L), turbulence = 1 (instead of 3), and fuel remaining = 40 (rather than 20). Within the particular constraints of this type of problem (four cases of acceptance and four cases of rejection), at most four partial rules are sufficient to specify all possible classifications. Wild cards are used to specify more general rules. For example (S,x,x) indicates a rule where size = S and turbulence and fuel remaining are unspecified. As each dimension has two possible values, such a partial rule describes four cases of acceptance and is sufficient by itself to classify all possible combinations of stimuli. Cases where a single rule is sufficient for classification correspond to the Category 1 classification category. Category 3 classifications require at least two partial rules to account for all possibilities. For example, (S,1,x) and (L,3,x) belongs in that category. Category 3 demands that all dimensions be specified and therefore requires at least four partial rules with no wild card. This representation

implicitly makes the Category 1 case easier to find and remember than Category 3, which should in turn be easier than Category 6. When being asked to classify a new triplet of stimulus values, if one of the current partial rules can match the stimulus then the result will be to accept, and to reject otherwise.

3.1.6.2 Learning Engine

The learning engine is responsible for the creation, modification, and deletion of the partial classification rules. It is based on the following principles of operation:

- Remove invalid rules – if there is a rule, with no wild card, that matches the stimulus and the feedback is negative.
- Specialize partial rules – if the feedback is negative and there is a rule with a wild card that matches the stimulus, then replace the wild card with the opposite value of the stimulus for this dimension.
- Create new simple partial rules (with two wild cards) – if the feedback is positive and if no rules currently cover the stimulus.
- Expand partial rule – if there is a rule where the first dimension matches the stimulus but the second is opposite and the feedback is positive, then create another complementary rule where the second dimension value matches the stimulus.
- Remove redundant rule – if a rule is a specialization of another more general rule; for example, (S,1,x) is a specialization of (S,x,x), then remove the more general rule.

These principles of operation are applied in this order when receiving feedback.

3.1.6.3 Metacognitive Strategy

One important aspect of the principles of operation of the learning engine is that most of them are dependent of the order in which the dimensions are evaluated. For example,

considering, size then turbulence and fuel (STF) does not provide the same result as considering fuel then size and turbulence (FST). The "create rule" principle produces (S,x,x) in the first case and (x,x,20) in the second. We designate the order in which the dimensions are evaluated as a strategy. In this study with three dimensions, six different strategies are possible. In general, no particular strategy is better than another, but for a specific classification problem, some strategies may be more efficient than others. For example, the strategy STF will immediately find the correct classification rule for a problem of type (S,x,x), but will struggle for a problem of type (x,x,20).

To remedy this problem, we use a meta-strategy to switch from one strategy to another: any time a strategy encounters N cases of negative feedback, another strategy is selected. To add a stochastic effect to the meta-strategy, N is randomly chosen, each time, between 1 and 3. Switching strategies might help some cases find the correct classification more rapidly, but it may also be disruptive because no particular strategy is allowed to run for an extended period.

3.1.6.4 Memory Moderation

Memory moderation was incorporated into the learning mechanism by applying the memory moderation algorithm on each trial prior to determining the classification response. First, forgetting from STM is applied, then transfer to LTM determined, then the classification response determined from all partial rules in LTM or STM.

The values for the parameters in the memory moderation algorithm were determined by using an initial value based on the Human Operator Simulator (HOS) work (Glenn et al., 1992; Lane et al., 1981), then adjusting those values based on some trial-and-error experimentation after the learning mechanism was integrated into the COGNET architecture. The value of T (the

length of the inter-trial interval) was determined from the task environment simulation to be 50 sec. The initial and final values for the learning parameters were:

a = STM decay parameter (initial value = .0003) (final value = 0.0015)
q_s = base probability for single-slot LTM conversion (initial and final value = .03)
b = LTM decay parameter (initial value = .2) (final value = 0.15)
t = time required to rehearse a single hypothesis slot (initial and final value = 0.5 secs.)

Using this memory moderation model, it takes more time to learn three-dimension rules than one- or two-dimension rules, since problem types are distinguished by the number of opportunities for simple solutions — one dimension rules, or two dimension rules in combination. Consequently, Category 1 and 3 problems are learned more easily than Category 6 because one- and two-dimension rules dominate while three-dimension rules are more likely to be discarded.

3.1.7 Model Tuning and Refinement

Model tuning involved refinements to response times on both the primary and secondary tasks and to probability of errors on the secondary tasks, to make the model performance more consistent with the tuning subject performance. Accomplishing these changes involved considering the effect of the primary task on the secondary task. Time consumption for motor actions and scanning behavior in the Phase I version of the AMBR model was accomplished using micro-models. Additional time consumption was added for rule application in the learning task. A workload coefficient was calculated based on perceived workload (across all tasks) and used as a scaling factor for secondary task response times. Probability of error on the secondary task was also adjusted, based on perceived workload. Response times for the primary task were adjusted by including two time charges:

- a response time charge based on number of strategy changes, and

- a response time charge for processing feedback based on whether the response was correct or incorrect.

3.1.8 Workload Model and Analysis

The Phase I effort developed a theory-based model of workload. The model was developed in three steps:

1) the six component TLX measures were conceptually mapped onto the internal architecture of COGNET;

2) the conceptual mapping was used to operationalize each measure in terms of specific aspects of metacognitive self-awareness, (whether they existed within COGNET or not);

3) those that did not already exist were then implemented and integrated into the COGNET metacognitive mechanisms. This resulted in the ability of the model to 'introspect' and create a post-hoc report of its own workload on the six measures involved.

The workload model is described in detail in Zachary, Le Mentec & Iordanov, 2001, and summarized in Table 1 below. In Table 1, the first column identifies a component measure calculated by the model, the second column identifies the computational formula for that measure, and the third column describes how the terms in the formula are operationalized in the COGNET architecture.

Measure	**Computational Formula**	**Operationalization**
PHYSICAL	$(\sum_{\text{all i}} k_i \bullet \textit{action-time}_i)/T_S$	Where i is an action type, k_i is the weight of that action type, and *action-time*$_i$ is a function which collects total time spent taking that action during a scenario by instrumenting the action mechanism
MENTAL	$(\sum_{\text{all i}} DPC_i \bullet DTT_i)/T_S$	Where i is a cognitive task, and DPC_i and DTT_i are available through metacognitive self-awareness
TEMPORAL	$-(\sum_{\text{all i}} \textit{nothing-to-do}_i)/T_S$	Where *nothing-to-do*$_i$ is a cognitive operator that returns a value of one *iff* the visual scan task has been completed

		and the model found nothing to do
PERFORM	$-(\sum_{\text{all i}} \textit{error-detected}_i)/T_S$	Where *error-detected*$_i$ is a cognitive operator that returns a value of one *iff* a cognitive task found that the model had made an error and was about to try to correct it (if correctable)
EFFORT	$(\sum_{\text{all i}} \textit{time-spent-in-work}_i)/T_S$	Where i is an increment in time, and *time-spent-in-work*$_i$ is a function measuring the time within that increment in which motor activity is occurring on any execution thread
FRUSTRATE	$(\sum_{\text{all i}} DTI_i)/T_S$	Where DTI_i is available through metacogitive self-awareness as a measure of the number of times cognitive task i was interrupted during the scenario

Table 1. Phase I Workload Self-Assessment Model

The workload self-assessment process added no parameters to the model or the underlying architecture. However, because the measurement scale of modeled processes were not the same as that used by the measurement instrument, i.e., the TLX quasi-interval scales, the measures generated by the model had to be calibrated to those specific measurement scales. This was done using a statistical regression, using data from human subjects used in the model-development training set. This calibration process added two regression values (slope and intercept) for each of the six TLX measures. The calibration formula used in Phase I can be found in Zachary, Santarelli, Ryder, Stokes & Scolaro (2000: Table 3-5).

It should be noted that the COGNET/iGEN model was the only AMBR model that produced a separate response for each of the six TLX measures as human subjects did.

Statistically, the Phase I model-generated workload assessments predicted the human values well in the Phase I training dataset, yielding r-squared statistics as shown in Table 2.

Workload measure	R^2 value
Physical:	36%
Mental:	38%
Temporal:	68%
Performance:	98%
Effort:	13%
Frustration:	53%

Table 2. Percent of Human Subject Variance (R^2) Accounted for by AMBR Phase I Model Predictions

3.1.8.1 Phase III Differences and Implications for the Phase 1 Model

In Phase III, there were three changes from Phase I:

- At the most superficial level, the moderator had changed the workload reporting scale from a 10 point to a 7 point scale.
- In addition, the non-learning (i.e., secondary) tasks themselves had been changed to create a single set of tasks that were different from either the unaided or aided tasks performed in Phase I.
- At a deeper level, the introduction of a learning task had required addition of a learning mechanism into COGNET itself.

The change in workload reporting scales required a simple recalibration of the mapping from model self-reports to the TLX reporting scales. This was done using the same statistical process used in Phase I. The change in tasks was analyzed as not relevant to the Phase I workload model. That is, it was concluded that the Phase I workload model should still be valid even with the changes to the ATC tasks, including the new learning (altitude change request) task.

The addition of a learning mechanism to the architecture, however, was more difficult. Given the change to the architecture itself, there was some question as to whether the workload self-assessment model created in Phase I would still apply. To assess this, an analog of the initial step in the Phase I workload modeling procedure was undertaken. Each individual TLX measure was analyzed to assess whether the self-assessment process it required might involve introspection into the (newly-added) learning mechanism. This analysis was inconclusive. On the one hand, it was deemed unlikely that measures such as physical workload or temporal workload would involve the learning process, because the attribute involved would be the same

regardless of the state of the learning process. The physical effort involved in responding to an altitude request, for example, was identical whether learning was in process or complete, and whether, during learning, the trial had a positive or negative outcome. On the other hand, other measures could depend on the state of the learning process. Mental workload, for example, might be assessed in terms of how difficult the learning process seemed. It was decided to initially apply the Phase I workload without changes other than the recalibration to the new seven point rating scales, to determine whether incorporation of learning mechanism features into the workload model would be necessary.

3.1.8.2 Application and Analysis of Phase I Model to Phase III Training Dataset

The initial application of the Phase I model to the Phase III human subject data in the training dataset yielded surprisingly poor results in terms of variance accounted for, as seen in Table 3. Particularly surprising was the fact that the physical demand and temporal demand measures were so poorly predicted, as these measures were predicted to be insensitive to addition of the learning mechanism.

Workload measure	R^2 value
Mental Demand:	0.03%
Physical Demand:	0.02%
Temporal Demand:	5.2%
Performance Err.	7.0%
Effort:	2.7%
Frustration:	1.5%

Table 3. Percent of Human Subject Variance (R^2) Accounted for by AMBR Phase I Model Applied to Phase III Training Data

To better understand why the predictions were so poor, we turned to an analysis of the scatter-plot relationships between model predictions and human data and found a surprising and disturbing set of results. All but one of the scatter-plots showed a general 'dot-stack' pattern, as shown in Figure 6 for the mental demand measure. The scatter plot shows that the model only

found three general levels of mental demand across all the trails. This can be seen in the fact that the model-generated workload assessments 'stack up' over three values on its self-assessment scale (the abscissa). This means that for all the different runs represented by data points at each of those abscissa values, the model found very little or no differences in mental demand. The fact that the values are spread across the ordinate, however, means that the human subject data did show variation over those same trials.

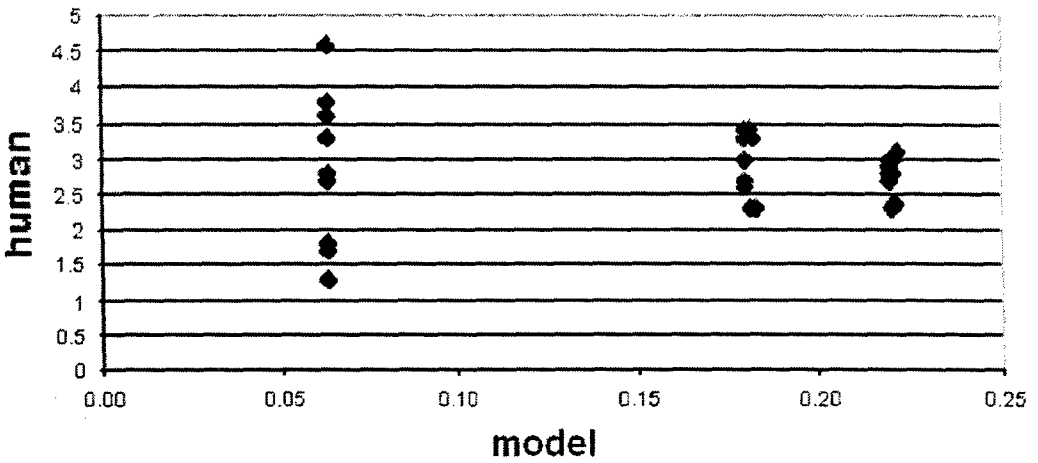

Figure 6. Scatter Plot for Phase III Mental Demand Workload Data Against Phase I Workload Model

Further analysis of the runs which comprise each stack of data points (i.e., 'dots' over one abscissa value) showed that they represented trials from a specific experimental condition -- control, dual-low, or dual-high. The trials with each dot stack thus varied only according to the category of learning involved. This means that the model was not generating any variability based on the category of learning, which was to be expected, because the Phase III learning mechanism was not instrumented or otherwise incorporated into the learning model. The implication from dot-stack plots like Figure 6, therefore, is that the human subjects were introducing variability based on the learning condition.

The same dot-stack pattern exhibited in Figure 6 was seen for all but one of the TLX workload measures. The performance measure was the one exception, although the 'dot-stacking' was less pronounced in the frustration measure as well. Figures 7 through 11 show the scatter plots for the other five workload measures, using the recalibrated Phase I model and the Phase III human subject training data.

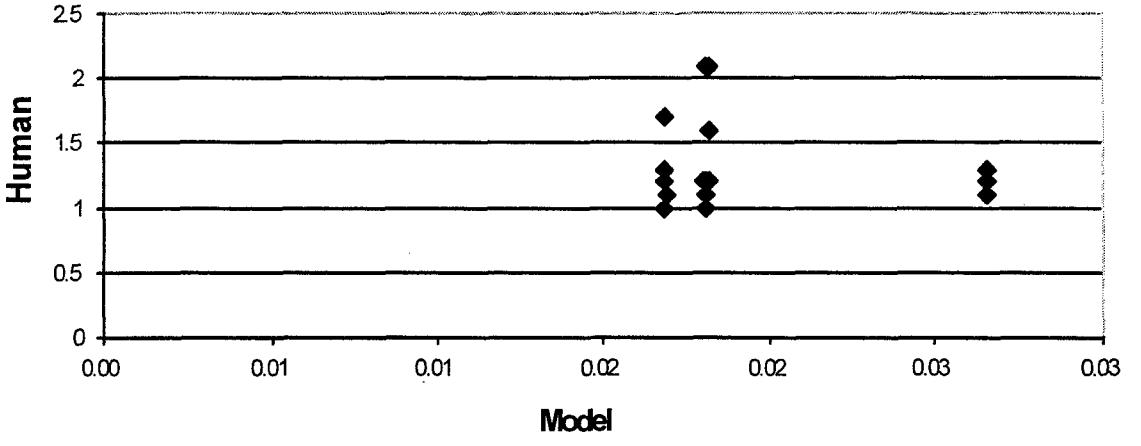

Figure 7. Scatter Plot for Phase III Physical Demand Workload Data Against Phase I Workload Model

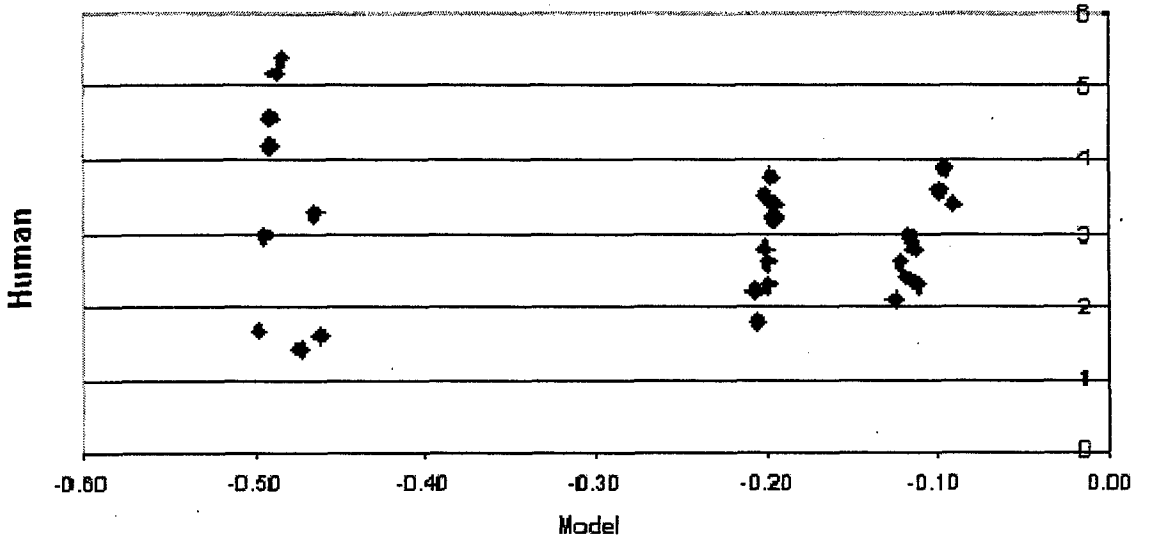

Figure 8. Scatter Plot for Phase III Effort Workload Data Against Phase I Workload Model

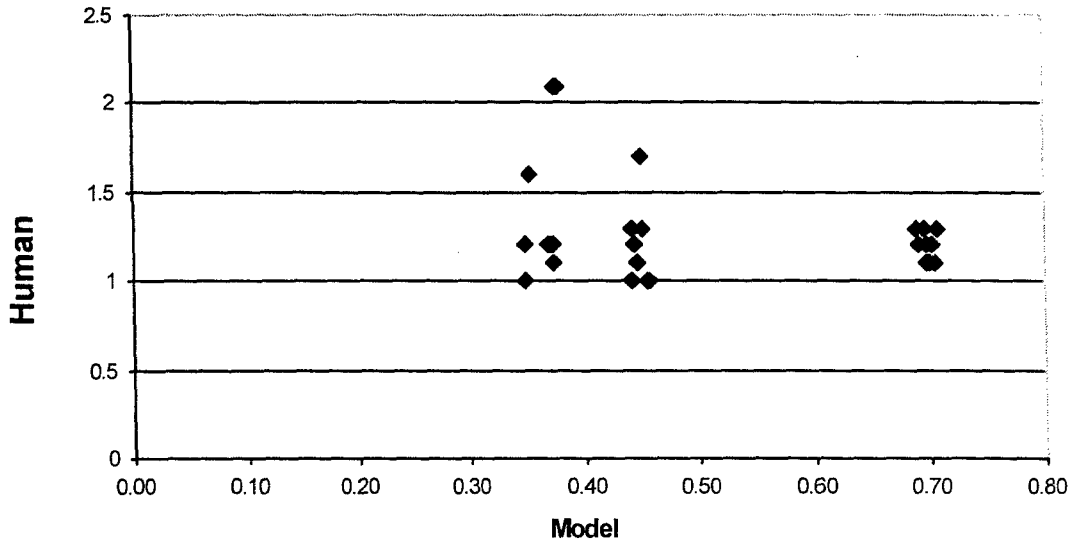

Figure 9. Scatter Plot for Phase III Temporal Demand Workload Data Against Phase I Workload Model

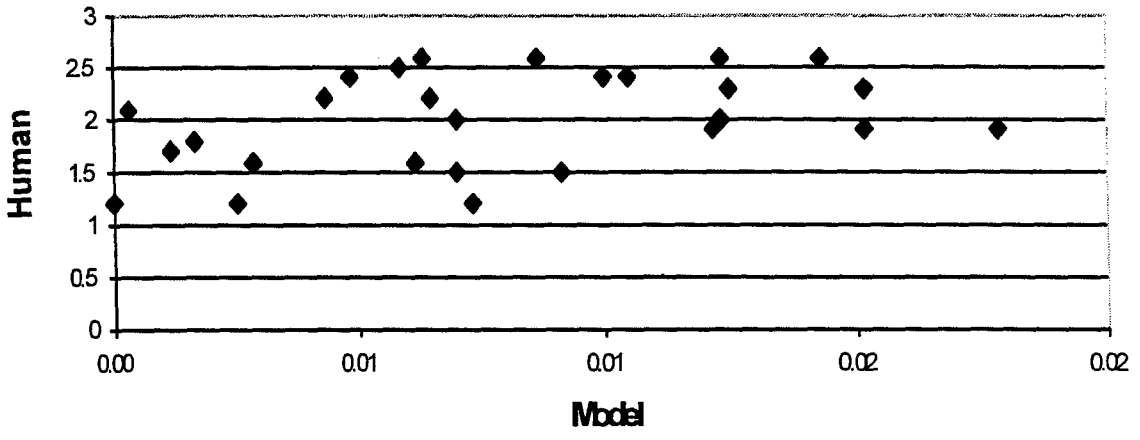

Figure 10. Scatter Plot for Phase III Performance Workload Data Against Phase I Workload Model

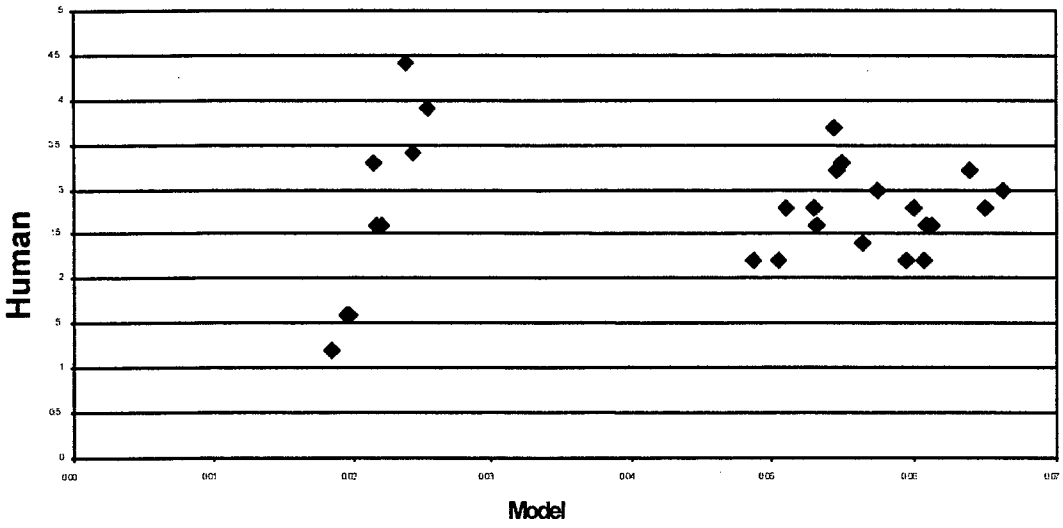

Figure 11. Scatter Plot for Phase III Frustration Workload Data Against Phase I Workload Model

3.1.8.3 Development of Phase III Workload Model

The data in Figures 6 to 11 were quite surprising, and opened up at least two major possibilities:

1) that the Phase I workload-assessment model had limitations which prevented it from mimicking human workload self-assessment in the Phase III tasks, and
2) that the human subjects were failing to follow the instructions of the workload assessment process, either consciously or unconsciously.

In the end, it was concluded that both possibilities were plausible, so the workload self-assessment process was modified accordingly.

The first possibility suggested that aspects of the learning mechanism did, in fact, have to be incorporated into the workload self-assessment process for measures that were logically learning sensitive -- mental demand, effort, performance and frustration. The method of doing this is discussed below. The second possibility suggested that the self-assessment process for

even those TLX measures which were *not* logically related to learning -- physical and temporal demand -- also had to be made to incorporate aspects of the learning mechanism. The would only be true if subjects were incorporating aspects of their learning experience during the just-completed trial into their workload self-assessment for physical demand and temporal demand, *even though the physical work involved and the time it required were totally independent of the learning aspects of the task.* Figures 7 and 9 clearly show that human subjects were varying their reports according to differences in learning condition. A variety of explanations for this fact are possible, but are generally out of the scope of this report. The simple fact that human subjects were reporting workload in this way was used as a basis for incorporating learning mechanism aspects into these two workload measures (physical and temporal demand) as well.

At the procedural level, the learning mechanism operated by developing a strategy for 'guessing' at rules, and then applying the strategy to produce and apply rules. A rule would be used until it led to an incorrect guess. After a number of incorrect guesses applying rules with a given strategy, the model would revise its strategy and begin again. The rules and past guesses were stored in a decaying memory. Thus, learning difficulties could be introspected as strategy changes. However, the number of strategy changes is directly related to the number of erroneous guesses, as it is only after a number of incorrect guesses that the strategy changes. Of course, the number of correct guesses is clearly related to the number of incorrect guesses as well (as the two sum to 16 in each trial). Given these relationships, it seemed that the number of incorrect guesses was the most clear aspect of the learning mechanism to incorporate into the workload self-assessment model, for two reasons: a) determinism rather than randomness; and b) clearer conceptual link to workload. Thus, the Phase 1 workload was modified in only one way, by

adding to each self-assessment rule, an added factor of the number of incorrect guesses in the just completed trial.

With the Phase I workload model adjusted in this single way, a new set of model predictions was generated and calibrated to the (new) seven point rating scale, and compared to the human subject training dataset. This Phase III workload model yielded strikingly better predictions as a result of the inclusion of the learning characteristics. The revised model was then incorporated into the overall AMBR Phase III model. The results of the model predictions are discussed below.

3.1.9 Model changes to handle transfer task

To handle the transfer task, the model was modified to map new values onto existing values and use the same mechanisms as in the learning task. New values were mapped to the most similar existing ones. Extrapolated stimuli were mapped 100% of the time to the nearest end value. Equidistant stimuli were mapped (arbitrarily) to the closest lower value.

3.2 Results

Analyses were performed by the moderator contractor, BBN Technologies. The Altitude Request learning task was the primary task, and the basic ATC task was the secondary task in the analyses. ANOVAS for Group (model vs. human), Category (Categories 1, 3, and 6) and Trial (trial blocks 1-8) were calculated for Primary and Secondary Task measures. Workload and Transfer ANOVAS were somewhat different and are reported in the relevant subsections below. Newman-Keuls tests were performed for significant main effects for some significant effects. Also, BBN calculated measures of variability (G^2 or SSE) comparing model and human results for many measures. The results reported here do not include any statistics; instead, they are

conclusions and discussions based on the BBN statistical analyses (ref to BBN AMBR Final Report). The figures shown in this report were also produced by BBN as part of their analysis.

3.2.1 Primary Task – Probability of Error

As shown in Figure 12, the model produced learning functions indistinguishable from human subjects on the primary task (the altitude change request task). Probability of error decreased over trials (trial effect) and was higher in the more difficult learning conditions (category effect) for both humans and model.

The solutions and intermediate results resemble human rule formulations. Category 3 was typically solved with either two 2-dimensional rules or one 2-dimensional rule and two rote cases (3-dimensional rules). Category 6 could only be solved by four rote cases and often was not solved by the end of eight blocks of trials. A typical model output at the end of the session was a near solution with three rote cases and one higher-level rule.

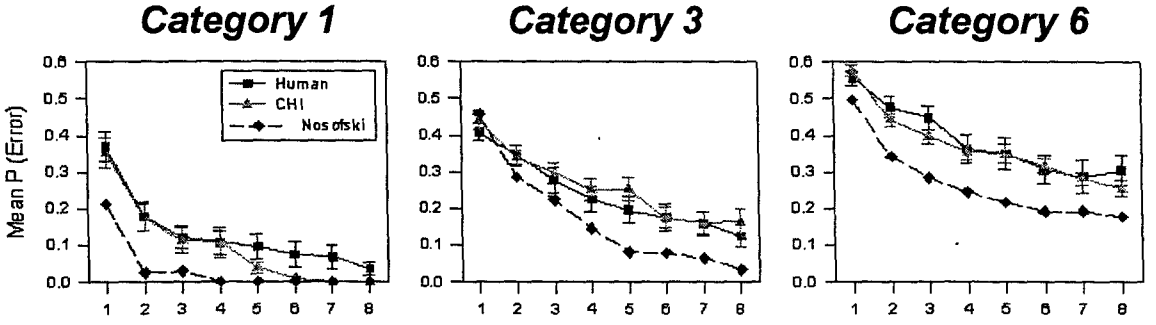

Figure 12. Probability of Error on Primary Task

3.2.2 Primary Task – Response Time

Figure 13 shows response time over trials for the primary task – the learning task. Humans showed both a category and trial effect, whereas the model only showed a category effect and no decrease in response time over trials.

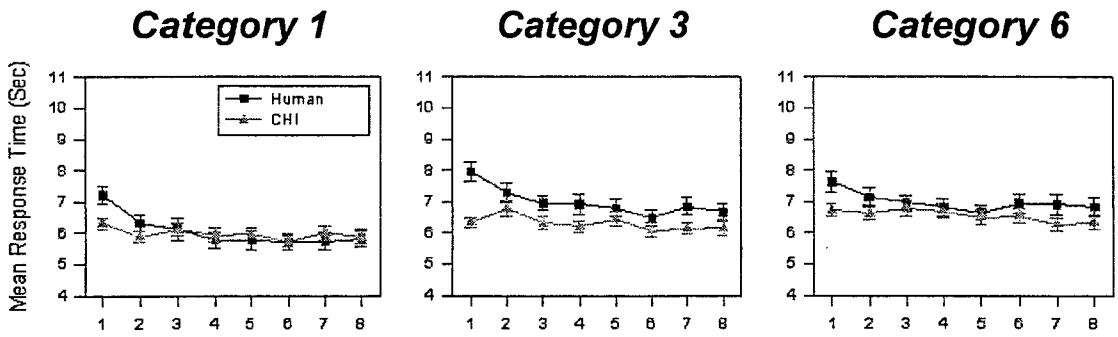

Figure 13. Response Time on Primary Task

3.2.3 Secondary Task – Penalty Score

For the secondary task – the hand-off task – neither the humans nor the model showed any significant effects. As Figure 14 shows, although there was some variability in the functions, there were no significant differences in secondary task due to the difficulty of the primary task.

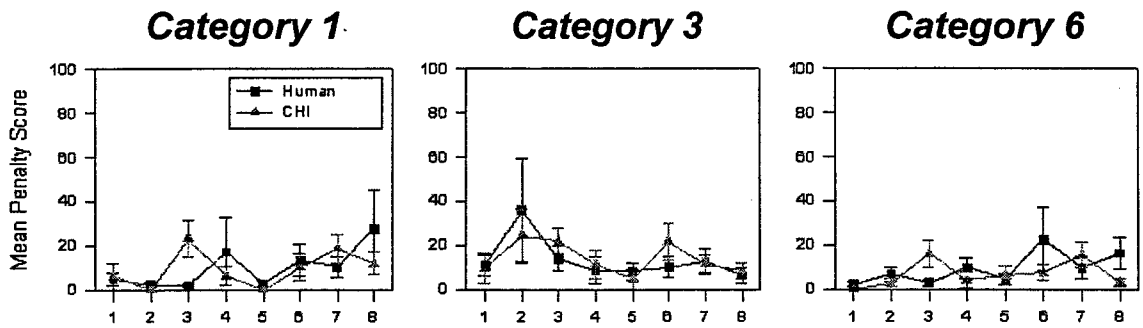

Figure 14. Penalty Scores for Secondary Task

3.2.4 Secondary Task – Response Time

As Figure 15 shows, the response times on the secondary task for the model were similar to those of humans. However, the model did not show the trial effect found for humans.

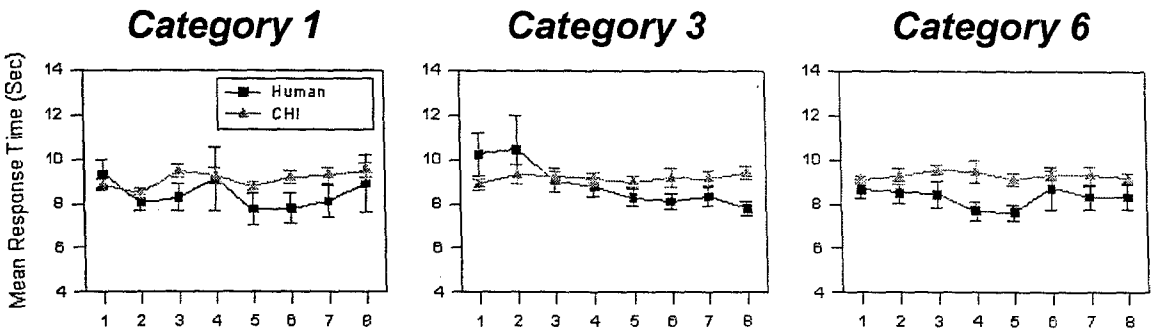

Figure 15. Response Time on Secondary Task

3.2.5 Workload Rating

For average subjective workload rating, the CHI model showed the same trends as humans. Figure 16 illustrates the increase in subjective workload as a function of learning task difficulty and decrease over trials for both human and model conditions. The consistently higher workload rating by the model was analyzed and determined to result from a mismatch in output format (1-based instead of the expected 0-based results).

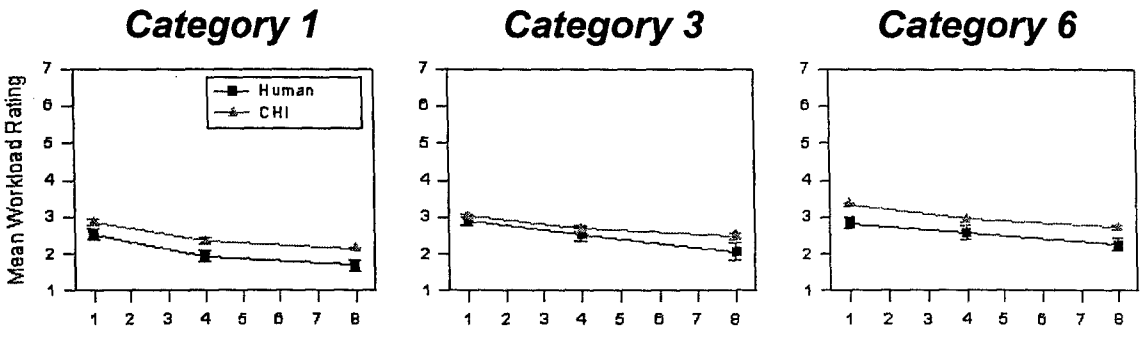

Figure 16. Workload Rating

3.2.6 Transfer Task – Probability of Error

The model predicted better transfer performance than humans actually exhibited, as indicated in Figure 17. The difference was greater on the extrapolated stimuli than the original 8 or trained stimuli. A reduction in model error for Trained and Extrapolated cases is possibly

related to cases of late solution by the model. If the model achieved a solution within block 8, then the model would improve its performance (lower probability of error) in the transfer block.

Humans exhibit some kind of interference or decay in both the Trained and Extrapolated cases, a result not seen in model performance. Thus, some factor involved in human performance is not reflected in the model. We reasoned at the time that we should be able to make the model show similar performance to humans by including some type of decay or interference to the transfer task in the model, and determined to address transfer performance in Phase IV.

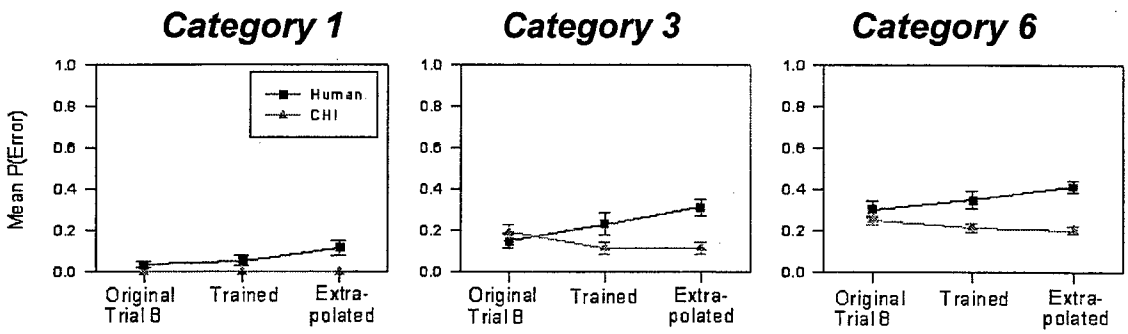

Figure 17. Probability of Error on Transfer Task

4. PHASE IV

Phase IV presented an opportunity, within a very limited budget, to improve, modify, and/or extend the Phase III model. We chose to focus on three main areas, as follows:

- improving the transfer task performance,
- correcting the workload assessment output, and
- incorporating a practice effect to improve accuracy of response times.

Each of these areas is described below.

4.1 Model Development

4.1.1 Confusion Factor in Transfer Task

Performance on the transfer task was modeled by mapping new stimulus values to the most similar ones in the original training set, with extrapolated stimuli always mapped to the end value closest to the new value, and equidistant stimuli mapped (arbitrarily) to the closest lower value. Initial comparisons with human data indicated a degradation in transfer performance by human subjects that was not accounted for in the model. Analysis of the human data indicated that performance was degraded overall due to the appearance of unfamiliar stimuli, and that the degradation was greater for untrained than trained stimuli. Thus, we incorporated a confusion factor (which could also be interpreted as a lack of confidence or cognitive bias effect) into the model. This was accomplished by reversing the trained response on a percentage of the transfer trials. The percentage reversed was smaller for "perfect learners" (those whose performance had reached asymptote on the final block of training trials. Since there were more perfect learners for Category 1 than Categories 3 or 6, there was a smaller transfer degradation for Category 1, matching human performance.

The algorithm used was the following:

When 0 errors in preceding trial block:
    Trained stimuli → use trained response 94% of time
    Untrained stimuli → use trained response 88% of time
When > 0 errors in preceding trial block:
    Trained stimuli → use trained response 76% of time
    Untrained stimuli → use trained response 64% of time

4.1.2 Workload Adjustment

The unit error noted in the Phase III results was corrected. 0-based values were output for analysis during the Phase IV runs.

4.1.3 Practice Effect

In addition, a practice effect (Lane, 1987) was added to the motor and scanning micro-models to account for the slight but steady decrease in time needed over trial blocks as a result of practice. The log of current scenario time, in seconds, multiplied by the log of trial number was computed [log(t)*log(trialNum)] and then scaled to approximate human secondary task performance [-x/95] and subtracted from action/scanning component of response time. This had the effect of creating the appropriate decrease in response time over trial blocks, but made the overall response time too low as the primary task response time was already shorter than human response time in Phase I results. Due to time constraints, this was not re-tuned during Phase IV.

4.2 Results

In the following subsections, as in Section 3.2, statistical analyses were conducted by the moderator contractor, BBN Technologies. Only discussion of the results is presented here. In the figures below, 'old' refers to Phase III data, while 'new' refers to Phase IV.

4.2.1 Primary Task – Probability of Error

No changes were made to the Phase III model, so the differences between Old and New are due to random variation. In both cases, there were significant category and trial effects for both our model and humans (see Figure 18).

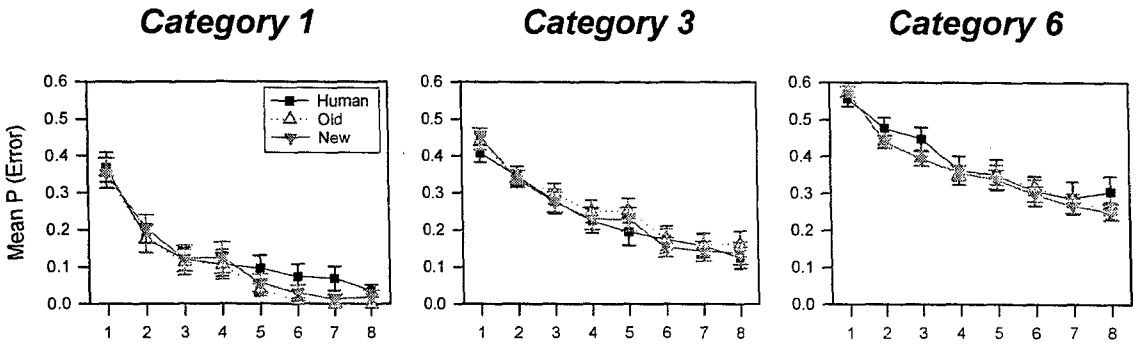

Figure 18. Probability of Error on Primary Task

4.2.2 Primary Task – Response Time

In this phase, our model showed a trial effect but no category effect, whereas the humans showed both effects (see Figure 19). Introducing the practice effect reduced average response time resulting in overall poorer approximation of human results, but introduced a trial effect not present in Phase III results.

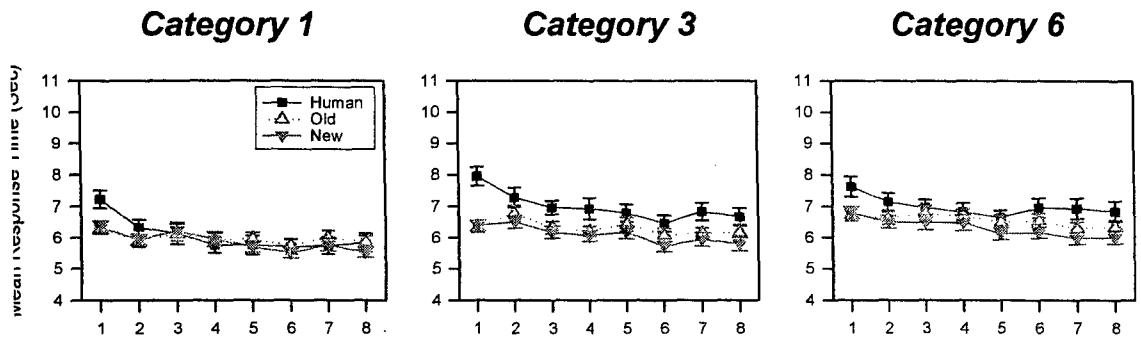

Figure 19. Response Time on Primary Task

4.2.3 Secondary Task – Penalty Score

Neither human subjects nor our model showed any significant effects (see Figure 20). Since there were no changes to the Phase III model, no changes would have been expected

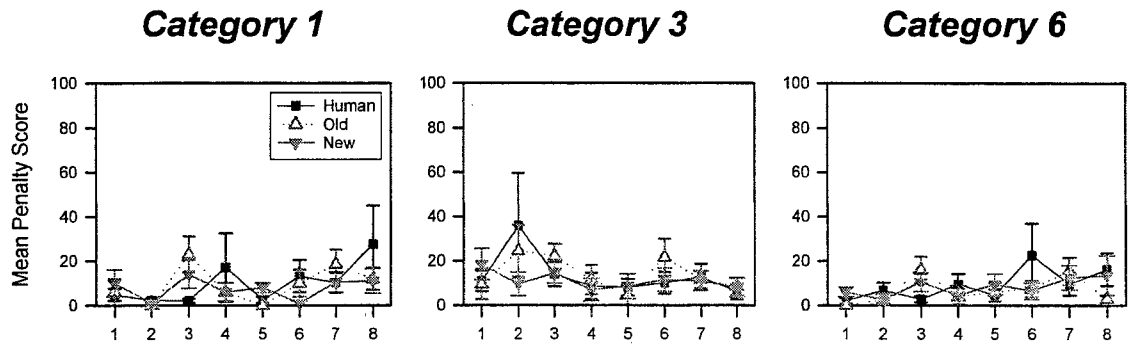

Figure 20. Penalty Scores for Secondary Task

4.2.4 Secondary Task – Response Time

Both humans and the model now show a trial effect, due to addition of the practice effect (see Figure 21).

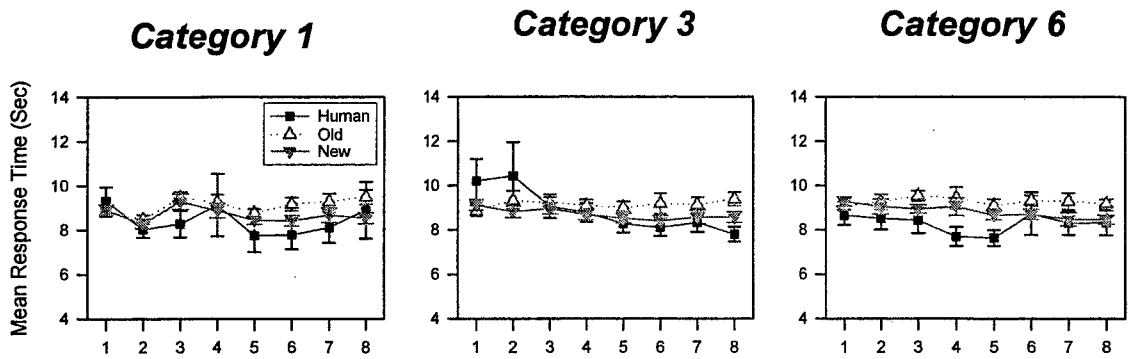

Figure 21. Response Time on Secondary Task

4.2.5 Workload Rating

As in Phase III the model provided workload rating trends very similar to humans. Both humans and the model showed significant trial and category effects but no significant interaction (see Figure 22). The consistently higher ratings generated by the model in Phase III were the result of an output formatting error. This was corrected in Phase IV, as noted above, providing a better overall approximation of human results.

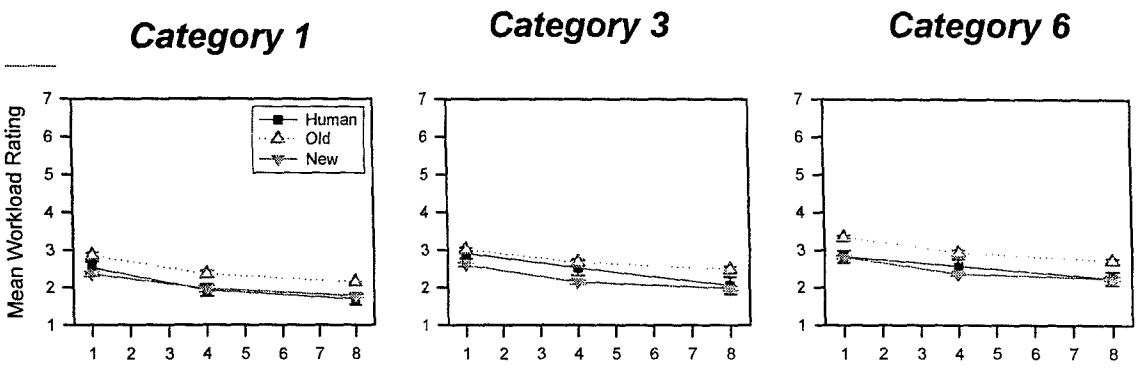

Figure 22. Workload Rating

4.2.6 Transfer Task – Probability of Error

The addition of the confusion factor in this phase degraded transfer task performance in a human-like manner (see Figure 23). Now, both humans and model had significant category and trial effects but no significant interaction.

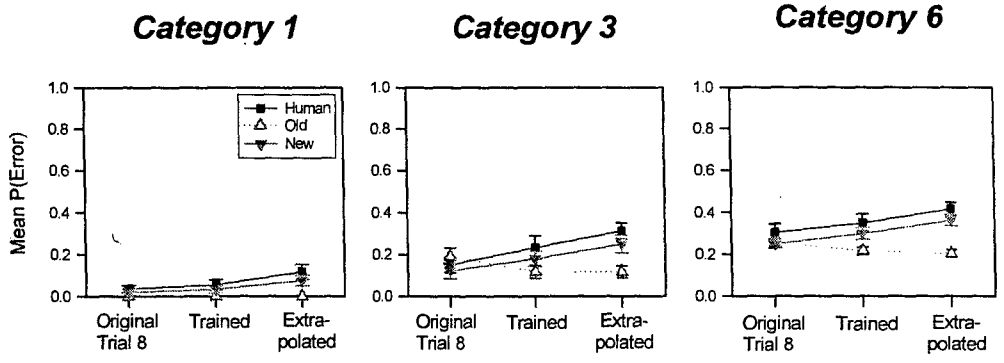

Figure 23. Probability of Error on Transfer Task

4.2.7 Probability of Error: Central vs. Peripheral

In Category III, theories of category learning predicted differential performance on central versus peripheral stimuli. Humans showed trial and category effects, but no interaction. As can be seen in Figure 24, the model showed significant trial and category effects and a significant interaction.

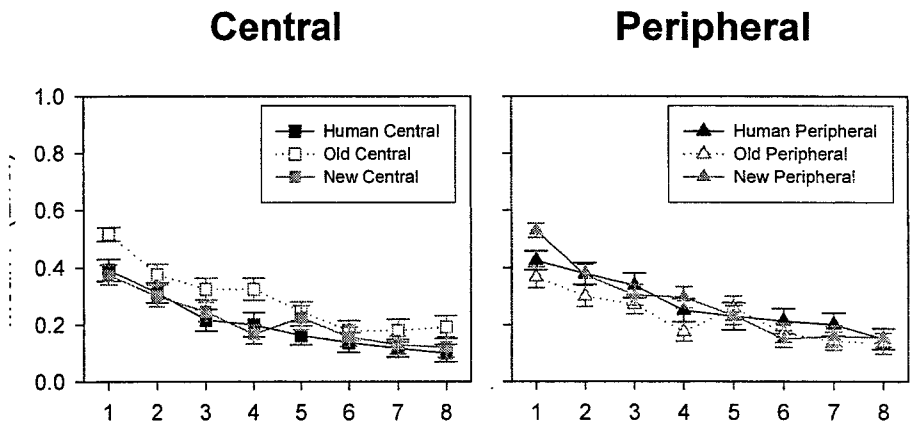

Figure 24. Probability of Error on Central vs. Peripheral Stimuli

4.3 Summary of Phase IV

The each of the three changes introduced in Phase IV improved model performance relative to human performance. The confusion factor added to the transfer task degraded performance in a human-like manner. The addition of the practice effect introduced the trial effect in response times that had not been present previously, and correction of workload output format improved approximation of human results on subjective workload judgments.

There was, however, one decrease in model performance due to the changes. The addition of the practice effect reduced average response time leading to a poorer overall approximation of human results for response time on the primary task. Although additional tuning might adjust the overall response time to be better in line with human results, we did not have enough time to make the modifications at the time.

5. PHASE IV – RESUBMIT

The opportunity to modify and resubmit the Phase IV models was provided. We decided to address the problem introduced in Phase IV regarding the primary task response time. At this

time it was also discovered that there were discrepancies in the human data used to develop the workload assessment model and this issue was also addressed.

5.1 Response Time

The results for primary task response time are shown in Figure 25. Overall response time is now closer to humans and the variability between model and human performance reduced.

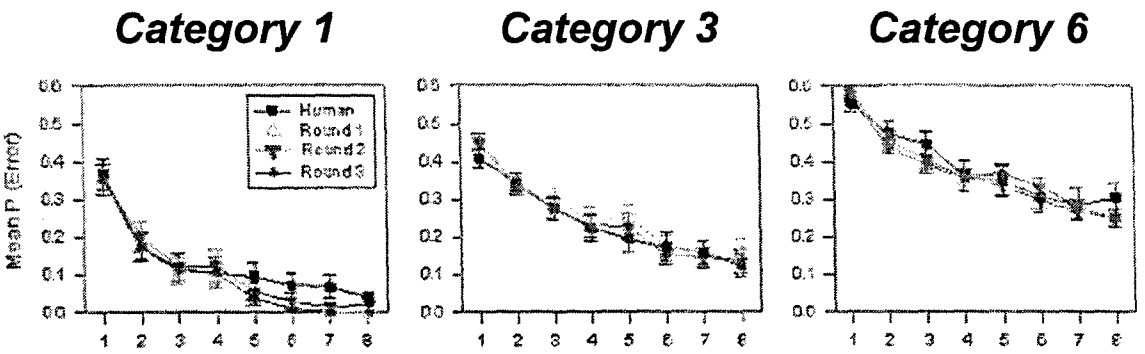

Figure 25. Response Time on Primary Task

5.2 Workload

The workload model development process described in Section 3.1.8 above was performed in Phase III. After the Phase IV model-human subject comparisons were made, during the documentation of the workload model, it was discovered that the raw human subject data on which the workload model was based had been partially corrupted in transit to the CHI Systems team. When this was discovered, a correct set of human subject data was obtained and the analysis was repeated to create a corrected Phase IV model. The calibration coefficients in that corrected model are given in Table 4. In Table 4, the terms in capital letters (e.g., PHYSICAL) refer to the component measures from the Phase I workload model (Table 1

above). The term PT refers to the number of incorrect guesses that the model made during the just-completed trial.

Calibration Formula
Physical demand = -10.5*PHYSICAL - .036* PT - 10.478
Mental demand = .08*MENTAL+ .126*PT + 2.308
Temporal demand = -1.57* TEMPORAL + .036* PT - 1.567
Performance = -87*PERFORM +.42 * PT + 1.91
Effort = .04*EFFORT + .16* PT + 2.1
Frustration = -.7*FRUSTRATE + .17*PT + 1.985

Table 4. Calibration Formulae for the Phase IV Workload Model Measures

This corrected Phase IV workload model yielded strikingly better predictions than the Phase I model (Table 3), as shown in Table 5, as a result of the inclusion of the learning characteristics.

Workload measure	R^2 value
Mental Demand	28%
Physical Demand:	15%
Temporal Demand:	30%
Performance Err.	58%
Effort:	41%
Frustration:	65%

Table 5. Percent of Phase III Human Subject Variance (R^2) Accounted for by AMBR Phase IV Workload Model

6. CONCLUSIONS

The research reported here added a learning capability to the COGNET/iGEN system, which previously had no such facility. Despite the limitations of the Shepard paradigm which motivated the larger project, learning was implemented as a general capability to learn the conditions under which each of a disjunctive set of goals or actions should be taken -- in other words, the ability to learn when to undertake different goals or actions. This approach was both

very general, and also consistent with the overall emphasis in COGNET/iGEN on procedural knowledge and multi-tasking. It was found that integrating this learning capability required other, deeper, extensions to the system, most notably the creation for the first time of an explicit short-term memory model and development of memory-performance moderators. These extensions, in turn, enabled the representation of memory decay, rehearsal, and proactive interference needed to model human learning performance. The manner in which the architectural changes were implemented allows future evolution of memory (including memory moderation) mechanisms and of learning mechanisms in COGNET/iGEN.

7. REFERENCES

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). *A Study of Thinking.* New York: Wiley.

Deutsch, S., & Benyo, B. (2001). The D-OMAR simulation environment for the AMBR experiments. *Proceedings of the Tenth Conference on Computer Generated Forces and Behavior Representation*, Norfolk, VA.

Ericsson, K., & Kinsch, W. (1995). Long-term working memory, *Psychological Review, 102(2)*, 211-245.

Glenn, F., Le Mentec, J., Ryder, J., Santarelli, T., Stokes, J., Zachary, W. (2003). Development of a Concept Learning Capability for a Human Performance Model. *12th Conference on Behavior Representation in Modeling and Simulation* (BRIMS '03), Scottsdale, AZ.

Glenn, F., Schwartz, S., & Ross, L. (1992). *Development of a Human Operator Simulator Version V (HOS-V): Design and Implementation.* Research Note 92-PERI-POX. Alexandria, VA: Army Research Institute.

Gluck, K., & Pew, R. (2001). Lessons Learned and Future directions for the AMBR Model comparison Project. In *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation.* Norfolk, VA, 113-121.

Lane, N., Strieb, M., Glenn, F., & Wherry, R. (1981). The Human Operator Simulator: An Overview. In J. Moraal and K.-F. Kraiss (eds.), *Manned Systems Design: Methods, Equipment, and Applications*. New York: Plenum Press.

Lane, N.E. (1987) Skill Acquisition Rates and Patterns: Issues and Training Implications. New York: Springer-Verlag.

Newell, A. (1990). *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

Nosofsky, R.M., Gluck, M.A., Palmeri, T.J., McKinley, S.C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: a replication of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition, 22*, 352-369.

Pew, R.W., & Mavor, A.S. (1998). *Modeling Human and Organizational Behavior: Application to Military Simulations*. Washington, D.C.: National Academy Press.

Pew, R.W., Tenney, Y.J., Deutsch, S., Spector, S., & Benyo, B. (2000). *Agent-based modeling and behavior representation (AMBR) evaluation of human performance models: Round 1 – Overview, task simulation, human data, and results. Final report*. Cambridge, MA: BBN Technologies.

Selfridge, O. (1959). Pandemonium: A paradigm for learning, in *Proceedings of the Symposium on the Mechanization of Thought Processes*, pp. 511-529.

Shepard, R.N., Hovland, C.L., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75* (13, Whole No. 517).

Tenney, Y.J., & Spector, S.L. (2001). Comparison of HBR models with human-in-the-loop performance in a simplified air traffic control simulation with and without HLA protocols: Task simulation, human data and results. *Proc. of the 10th Conference on Computer Generated Forces and Behavior Representation*, Norfolk, VA.

Waugh, N.C. & Norman, D.A. (1965). Primary memory. Psychological Review, 72, pp. 89-104.

Zachary, W. and Le Mentec, J.-C. (1999). A Framework for Developing Intelligent Agents Based on Human Information Processing Architecture. *Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing.* IASTED/Acta Press: Anaheim. pp. 427-431.

Zachary, W., Le Mentec, J.-C., and Iordanov, V. (2001) Generating Subjective Workload Self-Assessment from a Cognitive Model. In E. M. Altman, A. Cleermans, C.D. Schunn, & W.D. Gray (eds.). *Proceedings of Fourth International Conference on Cognitive Modeling.* Mahwah, N.J.: Lawrence Erlbaum Assoc. pp 229-234.

Zachary, W.W., Ryder, J.M., Ross, L., & Weiland, M.Z. (1992). Intelligent computer-human interaction in real-time multi-tasking process control and monitoring systems. In M.Helander and M. Nagamachi (Eds.), *Design for Manufacturability.* New York: Taylor and Francis.

Zachary, W., Ryder, J., Santarelli, T, & Weiland, M. (2000). Applications for executable cognitive models: A case-study approach. In *Proceedings of IEA2000/HFES2000* (pp. 1-737 to 1-740). Santa Monica, CA: Human Factors & Ergonomics Society.

Zachary, W., Santarelli, T., Ryder, J., Stokes, J., & Scolaro, D. (2000). *Developing a Multi-Tasking Cognitive Agent Using the COGNET/iGEN Integrative Architecture*, CHI Systems Technical Report 001004.9915. (Contract No. F33615-99-C-6007). Spring House, PA: CHI Systems, Inc.

NOTICES

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data, does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them

This report was cleared for public release by the Air Force Research Laboratory Wright Site Public Affairs Office (AFRL/WS) and is releasable to the National Technical Information Service (NTIS). It will be available to the general public, including foreign nationals.

Please do not request copies of this report from the Air Force Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161

Federal Government agencies and their contractors registered with the Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
8725 John J. Kingman Road, Suite 0944
Ft. Belvoir, VA 22060-6218

DISCLAIMER

This Technical Report is published as received and has not been edited by the Air Force Research Laboratory, Human Effectiveness Directorate.

TECHNICAL REVIEW AND APPROVAL

AFRL-HE-WP-TR-2005-0103

This technical report has been reviewed and is approved for publication.

FOR THE DIRECTOR

//SIGNED//

MARIS M. VIKMANIS
Chief, Warfighter Interface Division
Human Effectiveness Directorate